心理学报 2018, Vol. 50, No.8, 827−839
Acta Psychologica Sinica　DOI: 10.3724/SP.J.1041.2018.00827

联合评估和单独评估：富有潜力的助推手段

路 西[1, 2]　HSEE, Christopher K.[3]

([1] 北京大学光华管理学院，北京 100871) ([2] 中国农业大学经济管理学院，北京 100083)
([3] Booth School of Business, The University of Chicago, Chicago IL 60637, United States)

摘 要 人们有两种基本的评估模式：一是联合评估，即所有选项一起呈现和被评估；二是单独评估，各个选项单独地呈现和被评估。所有决策都是在其中一种评估模式或者两种模式的中间地带中做出的。联合和单独评估可以导致偏好反转和次优选择，理解其机制能帮助人们提高决策结果，包括慈善、定价、消费、甚至幸福感等方面。作为对各类决策影响广泛的重要变量，评估模式产生了不少反直觉的效应，目前研究尚不全面，是一块值得探索的沃土。

关键词 评估模式；偏好反转；广义可评估性理论；决策双系统；助推

分类号 B849:C91

人们的判断和决策总是在某种评估模式下作出的。现实状况和决策研究都表明，人们有两种基本的评估模式：联合评估(joint evaluation, JE)和单独评估(separate or single evaluation, SE)。在联合评估模式下，决策者同时面对两个或两个以上的选项；而在单独评估的模式下，决策者只面对一个单独选项(Hsee, 1996)。当然，单独评估和联合评估如同一个连续光谱的两端，现实生活中，人们有时会处于两种评估模式的中间地带。例如：为了决定报考志愿，一位高中生参观了几所大学校园。如果是数天之内密集地参观多所学校，那么他/她的评估模式就更接近联合评估。如果是隔几周或几个月才参观一所，那么他/她的评估模式就更接近单独评估，因为相较于当下的生动场景，许久前参观所获取的信息已经变得比较模糊不清。

单独评估和联合评估对于人们的决策过程和结果有着深刻影响，最为经典就是评估模式导致的偏好反转现象(Hsee, 1998; Kahneman & Ritov, 1994; Kogut & Ritov, 2005; List, 2002)。请想象你是一位音乐学院的学生，需要购买一本音乐类词典。词典A封面完好，内含1万词条；词典B封面破损，但内含2万词条。显然，当购买者同时看到两本词典时(即联合评估)，他/她对词典B的出价会更高。但是，如果单独评估，每个购买者只能看到其中一本词典，他/她对封面完好的词典A的出价更高 (Hsee, 1996)。

单独评估和联合评估的偏好反转(下文简称JE/SE偏好反转)至今仍然是研究的热点，各类相关现象丰富有趣，解释机制逐步细化，理论的延伸和现实的应用更是不断发展。

在新现象的发现层面，许多研究在探讨：单独评估中发现的效应是否在联合评估中仍然成立？还是会减弱、消失，甚至反转？例如 Chatterjee, Heath 和 Min (2009)探讨了单独评估中已经发现的心理账户效应是否在联合评估时仍然成立，发现只有当联合评估提供的情景间比较使得人们认为两种情况差异较小时(例如丢失 10 元的门票和 10 元现金在财务损失方面差异较小)，心理账户效应才会减弱。又如 Evers, Inbar, Blanken 和 Oosterwijk (2017)考察了同一个政策采用奖励框架(提高器官捐献者的受助优先级)或惩罚框架(降低非捐献者的受助优先级)的效果，发现单独评估时人们认为奖

收稿日期: 2017-08-13
两位作者在本文写作中贡献相同。
通信作者: HSEE, Christopher K., E-mail: christopher.hsee@chicagobooth.edu; 路西, E-mail: bewithlucy@gmail.com

励框架更合理，联合评估时，即使意识到两种框架的实际结果相同，人们仍然觉得奖励框架更合理，这种现象违反了规范性解释(normative explanation, 即理性解释)。

在理论解释层面，广义可评估性理论(general evaluability theory, GET; Hsee & Zhang, 2010; Zhang, 2015)从选项属性的可评估性(evaluability)的角度进行了分析，理论简洁，解释力最强。另外，想要和应要冲突(want-should conflict)、决策的双系统理论等也常用于解释 JE/SE 偏好反转，这些解释和可评估性也有一定的关联。

最后，更为重要的是，评估模式和广义可评估性理论可应用于政策制订，通过设计合理的评估模式，在对风险决策、跨期决策、慈善捐赠、消费选择、商业定价以及幸福感等方面助推人们做出更好的决策，提升决策的客观质量或者消费者的满意感受，甚至提高人们的幸福感。

本文将按照现象、内部机制、助推决策三个层次来介绍单独评估和联合评估的研究。首先，回顾 JE/SE 偏好反转的主要现象。其次，在解释机制层面，介绍可评估性和其他解释。最后，将评估模式作为助推手段，无论对消费者个人生活，还是公共领域(如慈善、健康等方面)，都能有所启迪。总之，单独评估与联合评估是一个既具理论重要性、又有前沿性的话题，是决策研究的蓝海之一，特别是在现实应用层面，如何设计恰当的评估模式来助推明智决策，评估模式的研究具有很高的潜力。

1 评估模式和偏好反转

1.1 JE/SE 偏好反转的主要现象

理性决策的恒定性原则假设，决策者不会受到备选方案表现形式的影响(von Neumann & Morgenstern, 1944)，但是关于评估模式的研究却发现，决策者违背了恒定性原则。相比于单独评估，联合评估时存在其他选项的比较，让人们产生偏好的变化，这类现象就称为 JE/SE 偏好反转。JE/SE 偏好反转在许多领域中都广泛存在。

在收入领域，Bazerman, Loewenstein 和 White (1992)考察了人们对两种报酬方式的偏好。方案 A 是自己获得 600 美元，他人获得 800 美元；方案 B 是自己和他人各获得 400 美元。在联合评估中，大部分人选择自己获利更多的 A 方案；但是在单独评估时，B 方案被评为吸引力更高。Hsee (1993)发现，当实验参与者同时面对“总收入较高但收入递减”或者“总收入较低但收入递增”这两种方案时，人们偏爱总收入较高的前者；但是，在单独评估中，收入递增的方案被认为更好。

在慈善捐赠中，Kahneman 和 Ritov (1994)发现，在联合评估时，面对两个慈善项目“海豚救助”和“为农民提供防晒以预防皮肤癌”时，人们对预防皮肤癌项目的捐款更多；但是，单独呈现一个慈善项目时，相比于皮肤癌项目，看到海豚救助时，参与者捐款更多。

在人事招聘中，Hsee (1996)请参与者想象自己作为咨询公司的老板需要招聘一位软件工程师。两位候选人都毕业于同一院校，其中候选人 A 的学业绩点(GPA)为 4.9 分(5 分制)，两年内写了 10 个程序；候选人 B 的 GPA 为 3.0 分(5 分制)，两年内写了 20 个程序。在联合评估时，人们愿意为候选人 B 支付更高的薪酬；但是在单独评估时，人们愿意为候选人 A 支付更高的薪酬。

在医疗决策中，Zikmund-Fisher, Fagerlin 和 Ubel (2004)让人们想象自己要预约医生为自己做眼部手术。医生 A 毕业于哈佛大学，进行过 80 次手术，都比较成功；医生 B 毕业于爱荷华大学，进行过 300 次手术，都比较成功。在单独评估时，人们预约哈佛毕业的医生 A 的意愿更强；但是在联合评估中，人们更倾向于预约手术经验较丰富的医生 B。

在道德判断领域，Paharia, Kassam, Greene 和 Bazerman (2009)研究了人们对 “间接引发较大危害的不道德行为”和“直接引发较小危害的不道德行为”的判断。在单独评估时，人们觉得直接引发危害的行为更不道德；但是在联合评估时，人们的道德判断发生反转，认为间接引发较大危害的行为更不道德。

总之，在产品消费、收入分配、人事招聘、医疗决策和道德判断等诸多领域，都存在着 JE/SE 偏好反转。值得注意的是，JE/SE 偏好反转不是由于因变量测量方式不同带来的虚假的回应差异。例如在音乐词典(Hsee, 1996)和慈善捐款(Kahneman & Ritov, 1994)等研究中，无论单独评估还是联合评估时，消费者的偏好都采用支付意愿(willingness to pay)来测量。接下来，我们将介绍 JE/SE 偏好反转中一部分特殊的现象，这类现象不仅和上述 JE/SE 偏好反转现象一样，违反了理性决策的恒定性原则，而且还违反了决策的占优性原则。

1.2 单独评估中违反占优性原则的现象

理性决策的占优性原则是指，个体永远都不会

采用一个被其他策略或选项占优的策略或选项。占优包括两种情况：一是强势占优，指一项策略或选项在所有方面都比其他策略和选项更好；二是弱势占优，指一项策略或选项至少在某一方面比其他策略或选项更好，在其他方面与其他策略和选项一样好(Von Neumann & Morgenstern, 1944)。评估模式的研究发现，在单独评估时，人们有时会违反占优性原则，偏好那些被其他选项占优的选项。

Hsee (1998)通过冰激凌实验演示了该现象。他请实验参与者想象自己在湖畔休息时想吃冰淇淋。单独评估时，参与者随机分配到以下两种情形之一(如图 1 所示)：摊位 H 出售 8 盎司冰激凌，装在容量 10 盎司的纸杯里；摊位 L 出售 7 盎司冰激凌，装在容量 5 盎司的纸杯里；联合评估时，参与者同时看到了两种情况。结果发现：联合评估时，人们对 8 盎司冰激凌给出了更高的价格；但是在单独评估时，人们的支付意愿违背了占优性原则，人们对溢出纸杯的 7 盎司冰激凌出价更高，虽然 8 盎司冰激凌才是占优选项。

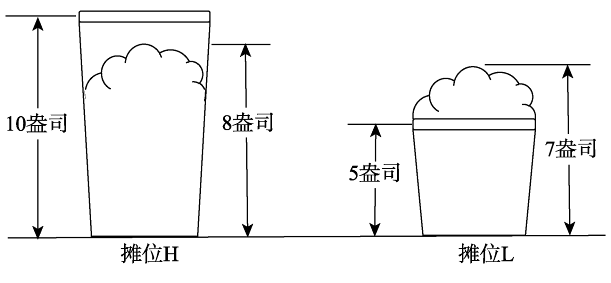


图 1　冰淇淋示意图
(资料来源: Hsee, 1998)

类似的，餐具套装的选择也是一个经典例证，2002 年诺贝尔经济学奖获得者卡尼曼教授在其获奖致辞中专门提及了此研究。假设你准备购买一套餐具，有两组套装。套装 A 共 24 件餐具，餐盘、碗和甜点盘各 8 个，每一件都完好无损。套装 B 共 40 件餐具，包含了套装 A 所有的 24 件，并且新增了杯子和酱料碟各 8 个，其中 7 个完好，9 个破损了。也就是说，除了完全一样的 24 件餐具，套装 B 还比套装 A 多出了 7 件完好的餐具，是客观上更好的选项。在联合评估时，人们对套装 B 的出价更高；但是在单独评估时，人们对套装 A 的出价更高(Hsee, 1998)。

这一效应在其他领域也得到了多次确证。在现实的二手棒球卡片交易市场中，List (2002)发现，即使是有经验的购买者也会受到此效应的影响。研究者展示了两套卡片：一套是 10 张崭新卡片套装；而另一套则是 13 张套装，包含了前一套装一样的 10 张新卡再加上 3 张中等旧的卡片。如果两套卡片摆放在一起，卡片收藏者对 13 张套装的出价更高。但是如果只摆出其中一套，卡片收藏者对 10 张套装的出价更高。此外，最新一项研究发现：在对他人学术能力的评估中，学术专家们(教授和博士生)也难免于此效应。有两位申请人的简历：申请人 A 的 8 篇论文都发表于排名靠前的期刊上，申请人 B 同样发表了 8 篇排名靠前期刊上的论文，但还额外有 6 篇论文发表在排名较后期刊上。如果专家们处于联合评估模式，同时看到两份简历，那么对两个申请人的学术能力的评估没有差异；但是当专家处于单独评估模式，只看到其中一份简历时，对申请人 A 的评估显著高于申请人 B (Powdthavee, Riyanto, & Knetsc, 2017)。

2　常见的解释机制

2.1　可评估性理论

为何在不同评估模式下，人们的偏好会出现反转？目前最简洁有力的解释是可评估性的影响。所谓可评估性，是人们对某个客观属性进行主观优劣判断时的难易程度和自信程度(Hsee, 1996; see Hsee & Zhang, 2010 as a review)。简单来说，属性的优劣越容易判断，则可评估性越高，越难判断，则可评估性越低。研究者对可评估性进行测量时，通常会询问参与者认为选项属性是否容易评估(e.g. Hsee, 1996; Hsee, Zhang, Wang, & Zhang, 2013)，例如大学生参与者认为 GPA 比编程经验更容易判断(Hsee, 1996)。

值得注意的是，在前文提到的所有 JE/SE 偏好反转中，所有选项总是涉及到两个属性的权衡——一个属性重要但难以单独评估；另一个属性不太重要但容易单独评估。所以，JE/SE 偏好反转的研究中，选项通常在两个属性上各有优劣(见表 1)。比如，在选择主刀医生的情境中 (Zikmund-Fisher et al., 2004)，手术经验(手术次数)是一个重要但是难以单

表 1　JE/SE 偏好反转的选项结构

选项	重要但难以单独评估的属性	不太重要但容易单独评估的属性
选项 A	较好 (例：完成 300 次手术)	较差 (例：毕业于爱荷华大学)
选项 B	较差 (例：完成 80 次手术)	较好 (例：毕业于哈佛大学)

独评估的属性；相对于经验，毕业院校却是一个没那么重要但是容易单独评估的属性。

在单独评估时，容易评估的属性对决策产生更大影响，即使该属性没那么重要；但在联合评估时，如果重要的属性变得更容易评估了，人们的偏好就会发生变化。仍以医生选择为例，单独评估时，人们更愿意预约哈佛毕业的医生，因为大学是一个容易评估的属性，即使人们知道手术经验更为重要，但由于没有参照点，人们难以判断进行过多少次手术算是经验丰富；在联合评估时，有两个选项互相对比，手术次数这个更重要的属性变得容易评估了，人们就会更愿意预约经验丰富的医生。这就是可评估性的变化会引起决策者的偏好反转。下面，我们将具体介绍哪些因素会导致可评估性的变化。

2.2 可评估性的影响因素

根据广义可评估性理论，某个属性的可评估性越高，人们对其价值的评估就越敏感，因为人们可以轻松自信地察觉到这一属性上的差异；而价值敏感性越高，该属性对决策的影响程度就更大。正是因为选项不同属性和可评估性的交互，JE/SE 偏好反转的现象才会发生。那么，什么因素会影响可评估性呢？

可评估性是依赖于参照系统的，存在参照信息的情况下，可评估性就较高。广义可评估性理论提出了三个重要的影响因素。(1)天性(nature)，是内在的参照信息，反映了人类长期进化而来的参照信息体系；(2)知识(knowledge)，是习得的参照信息，即过去是否学习和积累了目标属性的数值分布作为参照信息；(3)评估模式(mode)，是当下的参照信息，是否有其他选项作为参照信息。只要有一项因素的可评估性较高，可评估性就高，相应的价值敏感性也越高(如图 2 所示)。

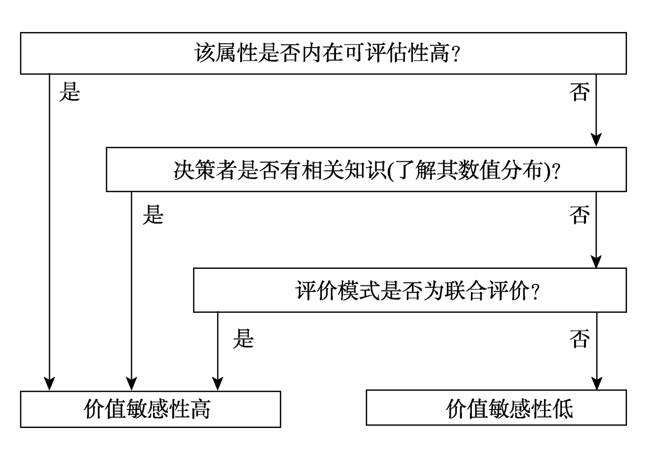


图 2 可评估性和价值敏感性的影响因素

因素 1：天性，是指人类对于所评估的价值是否有天然存在的、稳定的生理或者心理度量(即一个参照系统)。如：舒适温度、睡眠充足与否、社交孤独等方面，人们无需学习，即有天然的、内在的参照系统可以做出评估。但是，对于钻石的大小、汽车的马力等问题，人们没有天然的、内在的评估系统，需要学习相关知识或者比较其他选项才能评估其优劣。因素 2：知识，是指评估者本人是否了解目标属性的数值分布信息(例如：数值范围、均值等)。例如：经常购买红酒的消费者对红酒的年份、产地等信息拥有更多知识，在出价的时候比新手更依赖于知识进行判断(Gustafson, Lybbert, & Sumner, 2016)。一般而言，人们对于类别信息的知识较多，但是对于数量信息的知识较少。比如：我们都知道收益是好的，损失是坏的，但是收益多少钱、收益率有多高才是好的，不是理财专家就难以评估。因素 3：评估模式，也就是单独评估还是联合评估，在单独评估时，因为不存在选项之间的参照，可评估性较低；联合评估时，由于选项可以互为参照，可评估性较高。

基于广义可评估性理论，JE/SE 的偏好反转现象其实可以归为两种类型。一类是评估模式和选项属性(天性)的交互作用，另一类是评估模式和选项属性(知识)的交互作用。

2.2.1 评估模式和选项属性(天性)的交互

评估模式和选项属性(天性)的交互作用中，选项一般涉及到两个属性，其中一个属性是不太重要但是内在易评估的(例如电子相框是否美观)，另一个属性是更为重要但内在难评估的(例如电子相框分辨率的高低)，需要通过比较的方式来评估的。相比于内在难评估的属性，内在易评估的属性在单独评估中权重更大；但是到联合评估的模式下，因为有选项间的参照，内在难评估的属性可评估性提高，在决策中的权重也会变大。Yang, Hsee, Liu 和 Zhang (2011)让人们选择两个电子相框：一个分辨率更高但是外观难看，一个分辨率低但是外观漂亮。当处于单独评估时，人们更多选择好看的相框，因为美观是一种内在可评估性很高的因素，我们无需比较就可以判断美丑；当处于联合评估时，人们更多选择难看但分辨率高的相框，因为通过选项间的互相比较，分辨率这一内在难以评估的属性变得更容易评估了，人们的偏好就发生了变化。值得注意的是，在本研究中，研究者对单独评估组也采用了组内设计，参与者首先观看一个电子相框，然后有 10 分钟

的间隔时间完成填充任务，然后再观察另一个相框。这部分研究表明，单独评估和联合评估的差异不是组间和组内设计的差异导致的，而确实是评估模式的差异造成的。

2.2.2　评估模式和选项属性(知识)的交互

评估模式和选项属性(知识)的交互作用中，选项一般涉及到两个属性，其中一个属性是不重要但是人们知识较为丰富的，另一个属性是重要但是人们的知识较为缺乏的。因此，不重要但是知识丰富的属性在单独评估中权重更大，重要但是知识缺乏的属性在联合评估中权重更高。例如，医疗决策的研究发现，单独评估时，人们更偏好成功率 33%但距离自己家 15 分钟的生育门诊，但是在联合评估时，人们更偏好成功率 40%但距离自己家 40 分钟的生育门诊(Zikmund-Fisher et al, 2004)。因为花费时间这一维度上，人们拥有更多的知识和了解；但是医院的成功率这一维度上，人们的了解较少，在联合评估时有选项间的参照才能提高其可评估性。

评估模式和知识的交互中存在一个典型子类别，即评估模式和类别−数量信息的交互。以前文提到的餐具套装的选择为例(Hsee, 1998)，一套餐具数量多但有破损，一套餐具数量少但都很完好，选项中存在餐具数量和是否完好两个维度。在单独评估时，餐具是否完好作为类别信息，人们拥有较多了解，容易评估，因此餐具是否完好对评估的影响高；而在联合评估时，因为有了选项间的参照，数量信息的可评估性提高了，对决策的影响也增大了。类似原理的，关于道德判断的研究中(Paharia et al., 2009)，研究者对比了两种不道德行为(间接引发较大危害的行为和直接引发较小危害的行为)。在单独评估时，直接还是间接引发是类别信息，不需要参照也很容易评估，因此人们觉得直接引发危害的行为更不道德；但是在联合评估时，危害程度这种数量信息的可评估性提高了，因此人们对两种行为的判断反转了。

2.3　其他解释

虽然广义可评估性理论是目前解释力最强的理论，但还有其他解释机制提供了不同的视角。下面我们将简要予以介绍。

Bazerman 等(1992)提出了 “想要和应要冲突” (want-should conflict)的解释，他们认为，在单独评估时个体更多考虑到“想要” (want)的因素，但是联合评估时个体更多思考“应要” (should)的因素。例如：Okada (2005)发现，面对享乐选项(50 元餐厅代金券)和实用选项(50 元超市食品代金券)的取舍，在单独评估时，人们更喜爱享乐选项，在联合评估时，人们更喜欢实用选项。类似的，Shaffer 和 Arkes (2009)发现，人们在联合评估中更喜欢现金报酬，但是在单独评估时，非现金报酬(特别是享乐品而非实用品)更受人喜欢。这类现象可以用想要和应要的冲突来解释。单独评估时，享乐选项激发了更多想要的渴望，而联合评估时，如果要选择享乐选项，人们会考虑应要的因素，比如是否存在合理的理由。不过，这类现象也可以用可评估性解释，相比于实用性，享乐属性的内在可评估性较高，因此在单独评估时影响更大。

Kahneman 和 Ritov (1994)提出了组内比较和组外比较的差异。他们探讨了“救助海豚项目”和“预防皮肤癌项目”偏好反转的原因，指出：在单独评估时，人们更多进行组内比较，海豚组把海豚救援和其他动物救援比较，海豚属于动物中人们更愿意救助的；防晒组把增加防晒措施来预防皮肤癌项目和其他人类救援比较，这一项目属于人类救援中比较不重要的。联合评估时，人们更多进行了组外比较，显然救助人类比救助动物更重要。其实组内组外比较的核心机制和可评估性一样，都是参照信息。用广义可评估性理论来解释，单独评估时，救海豚和救人两个项目内在的参照系统是不同的；但是联合评估时，当前的参照选项就成为更凸显的参照点，参照点的变化引起了偏好改变。

最近，Sher 和 McKenzie (2014)提出了“选项即信息”模型(option as information model)，指出：个体在看到选项前对属性的数值分布有一个推断，但每看到一个选项就类似一次抽样，每次抽样后都会更新自己推断的模型，再根据新形成的参照模型来判断。从本质上而言，选项即信息模型和可评估性都采用了参照信息来解释，但该模型的重点在于论证 JE/SE 偏好反转并不是非理性的，反而是因为有了其他选项作为新信息，人们理性地重新推理，才发生了偏好的改变。

此外，双系统理论也常被用于解释 JE/SE 差异。在单独评估中，个体更多依赖系统 1 的启发式加工；在联合评估中，个体更多依赖系统 2 的分析式加工(Gino, Moore, & Bazerman, 2008; Bazerman, Gino, Shu, & Tsay, 2011)。Gino 等(2008)提出，日常生活中，我们可能面临两类不道德行为：一种是过程重度不道德，恰巧有积极后果，另一种是过程轻度不道德，恰巧有消极结果(所谓恰巧，是指行为

者并不能控制后果，结果好坏只是随机概率问题)。在单独评估时，人们容易出现结果偏差，觉得“过程轻度不道德但恰好产生消极结果的行为”更不道德；在联合评估时，偏差会减弱。为了证明双系统的影响，研究者在单独评估时，启动参与者进行直觉或者理性加工，发现理性加工确实能减弱结果偏差。但是，值得注意的是，如果该研究能在联合评估的模式下启动系统 1 的加工方式也发现结果偏差增大，才能更确切地证明其中的因果关系。同样的，这一现象也可以用可评估性来解释，结果的积极或消极是一种类别信息，在单独评估时可评估性较高；而过程的轻微或重度不道德是数量信息，在联合评估时可评估性较高。

广义可评估性理论中的评估模式和选项维度(知识或天性)的交互与双系统有一定程度的相似。一般而言，内在可评估性高的因素经常和系统 1 比较相关，依赖于直觉和情绪，内在可评估性低的因素和系统 2 相关，依赖于分析和计算。正如前文提到的外观美丑，是一种内在可评估性很高、且依靠直觉判断的因素；而各类数字信息，无论是手术经验、屏幕分辨率的高低等等，都是需要分析、计算才能理解的。类似的，知识丰富的信息，可能已经转化为个体自己的经验法则，依靠系统 1 的启发式加工即可迅速完成；而知识贫乏的信息，如果人们不进行较慢的分析和计算则难以推断和评估。

虽然双系统理论具有很高的普适性，但是，在 JE/SE 偏好反转问题上，广义可评估性理论具有更为独特的解释力，许多现象单纯使用双系统理论难以充分解释。例如：Zikmund-Fisher 等(2004)关于医生选择的研究，在单独评估时，人们更愿意选择毕业于哈佛大学、进行过 80 次手术的医生，在联合评估时，人们更愿意选择毕业于爱荷华大学、进行过 300 次手术的医生。即使在单独评估时要求人们理性思考采用系统 2 加工，或者在联合评估时增加人们的认知负荷使人们采用系统 1 加工，恐怕仍然难以消除这种偏好反转，并不是只要采用系统 2 的分析思考就能完全弥补知识缺乏或者内在难以评估带来的影响。

3 评估模式助推明智决策

研究人类的判断和决策，不仅是希望解释其中的规律，更重要的是去影响现实，帮助人们改善和提升决策的结果。单独评估和联合评估在现实中可操作性极高，因此，完全可以通过对决策模式或信息呈现的设计，助推人们的决策。下面，我们将从以下几个不同的领域，分别介绍评估模式会如何影响人们的决策，希望能启发政策的设计者选择合适的评估模式，达成助推的目的。

3.1 风险决策

生活在充满风险和不确定性的世界中，个体的许多经济决策都与此相关，例如理财投资中更保守还是更冒险，是否购买彩票奖券等产品。Hsee 及其同事提出：评估模式对于个体的风险决策有显著影响(Hsee, Zhang, Wang, et al., 2013)。

众所周知，在风险和跨期决策的研究中，最重要的就是三种函数：一是效用函数(value function or utility function, Kahneman & Tversky, 1979)，二是时间折扣函数(temporal discounting function, Fredrick, Loewenstein, & O'Donoghue, 2002)，三是概率权重函数(probability weighting function, Kahneman & Tversky, 1979)。这三个函数都是非线性的，特别是在原点附近(即价值为零、延期程度为零、概率为零或百分百的时候)更加陡峭。单独评估和联合评估可以影响函数的非线性程度(如图 3 所示)，在单独评估的模式下，函数的非线性程度较高；在联合评估的模式下，函数的非线性程度较低(Hsee & Zhang, 2010; Hsee, Zhang, Wang, et al., 2013)。基于广义可评估理论，在单独评估时，人们对于类别的变化更为敏感，例如价值从零变为正数时、风险从无风险变为有风险时，都是类别的变化，因此人们的效用判断发生了很大改变。但若无其他的数量信息作参考时，人们对数量的变化较为不敏感，例如发生概率从 10%变为 20%时，人们对数量变化的效用判断变化就不太大。因此，和联合评估相比，单独评估时的效用函数、时间折扣函数和概率权重函数在原点附近的非线性程度都更高。

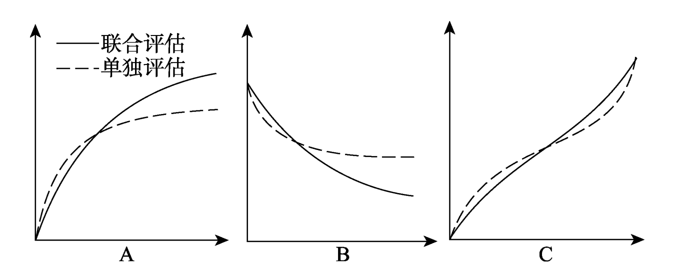


图 3 评估模式和效用函数(A)、时间折扣函数(B)、概率权重函数(C)
(修改自: Hsee & Zhang, 2010)

现在，首先讨论风险决策领域(查看图 3 中的概率权重函数)，在确定性点附近(概率为 0%或

100%), 单独或联合评估时, 人们的敏感程度是类似的; 但是, 随着风险水平的上升, 单独评估的情况下, 人们对风险的变化更不敏感。例如, Hsee 等(2013)发现在公共政策方面, 假设有两种应对洪水的方案, 方案 A 可以 100%地保护 2 万人安全(无风险、小收益), 方案 B 有 90%的概率可以保护 4 万人安全, 但是有 10%的概率保护不了任何人(有风险、大收益)。相比于联合评估, 单独评估时参与者对方案 A 的偏好程度更高。更近一步, 研究者们请大学生考虑附近新开的百货商场正在赠送价值 50 美元的开业礼券, 获奖概率分别为 100%, 90%和 80%, 假设自己有 120 分钟的空闲时间可以去排队, 问他们最多愿意等多少分钟。在联合评估中, 获奖概率 100%, 90%和 80%这三种条件下, 参与者愿意等待的时间逐渐下降, 并且两两之间差异显著。在单独评估中, 对比获奖概率 100%和 90%的两组, 人们愿意等待的时间有显著差别; 但是, 对比获奖概率 90%和 80%的两组, 人们愿意等待的时间并没有显著差异, 和联合评估的结果不同。这就是说在单独评估时, 人们只对于有无风险这一类别信息敏感, 对于多大程度的风险(例如获奖概率 90%还是 80%)这一数量信息不敏感。

此外, 关于不确定性情况下, 评估模式如何影响决策还缺乏系统深入的研究。早期 Fox 和 Tversky (1995)发现了比较忽略效应(comparative ignorance), 认为人们的不确定性规避其实来源于和确定性的备选方案的比较, 而当不存在明显的比较信息时, 人们其实并没有那么回避不确定性。他们邀请参与者想象自己面对这两个游戏袋, 里面装了红色和黑色筹码, 只要抽到红色筹码就可以获得 100 美元。袋子 A 中有 50 个红色筹码, 50 个黑色筹码; 袋子 B 中红色和黑色筹码一共 100 个, 但是具体每种多少不清楚。研究者询问了参与者愿意出价多少来参加这个游戏。在联合评估时, 人们对于袋子 A 的出价显著高于袋子 B, 表现出了不确定性规避; 但是在单独评估时, 人们对两个袋子的出价却没有差异。这和 Hsee, Zhang, Wang 等(2013)的研究中, 人们在联合评估时才会在意风险的大小(对概率数值敏感)其实是比较符合的。单独评估时, 人们对袋子 A 和 B 都感到结果无法确定, 却区分不出哪个不确定性更高, 只有联合评估时, 才知道 B 比 A 更难以预测。

风险决策中的评估模式差异其实是 JE/SE 偏好反转的一个子类, 只是处于这种特定的情境下。其根源还是选项存在两个不同的属性: 一个属性为类别属性——有无风险; 另一个属性为数量属性——风险大小。在单独评估时, 有无风险的可评估性更高, 人们对此较为敏感, 风险大小的可评估性较低, 人们对此较为不敏感; 但是, 在联合评估时, 人们对两类属性都较为敏感。

因此, 在风险相关的决策中, 如果政策的制定者希望消费者能区分出风险程度的高低, 对概率的数值变化更加敏感, 那么就应该使消费者处于联合评估的模式中, 例如提供股票和债券两种投资方式的损失概率对比, 这样消费者就会对概率的数值更敏感。如果政策制定者希望消费者对有无风险这一类别信息更为敏感, 那么单独评估的模式更为有效。

3.2　跨期决策

正如前文已经提及的, 在跨期决策领域(观察图 3 中的时间折扣函数), 在零点附近(没有延迟), 单独或联合评估时, 人们对时间折扣的敏感程度差不多的; 但是, 当延迟时间增加后, 单独评估时, 人们对时间变化没有联合评估时敏感(Hsee, Zhang, Wang, et al., 2013)。例如, 单独评估下, 人们还会表现出更强的现时偏差(present bias, Laibson, 1997)。让人们考虑两种新的网络服务套餐: 套餐 A 速度 2MB 每秒, 现在即可升级; 套餐 B 速度 4 MB 每秒, 但是 3 个月后才可以升级。单独评估时, 参与者对套餐 A 的偏好程度比联合评估下更高(Hsee, Zhang, Wang, et al., 2013)。此外, 进一步研究发现, 让大学生考虑他们为一个初中生补一次课, 就可赚取 100 元报酬, 如果报酬立刻获得、1 个月后获得、3 个月后获得, 他们愿意付出的补课时间为多久(在 0~240 分钟之间)。在联合评估中, 立刻、1 个月后、3 个月后获得报酬这三种条件下, 大学生愿意付出的补课时间递减, 并且两两差异显著。但是, 在单独评估时, 立刻和 1 个月后获得报酬的两组, 大学生愿意付出的补课时间有显著差异, 但是 1 个月后和 3 个月后获得报酬这两组之间无差异; 也就是说, 单独评估时, 人们对于 1 个月还是 3 个月延迟的数量信息并不敏感, 人们只对是否有延迟这一类别信息敏感。

和风险决策类似, 跨期决策中的评估模式差异也是 JE/SE 偏好反转的一个子类别。在跨期决策情境中, 选项的类别属性为有无延迟, 数量属性为时间长短。在单独评估时, 有无延迟的可评估性更高, 人们对此较为敏感, 时间长短的可评估性较低, 人们对此较为不敏感; 但是, 在联合评估时, 人们对两类信息都较为敏感。

因此，在涉及到跨期决策时，如果期待人们更耐心、更愿意延迟获得较大奖赏，那么政策制订者应该更多让人们采用联合评估的模式；但是如果需要鼓励人们获得即时满足，那么政策制定者可以设计单独评估的模式。

3.3 慈善捐赠

许多关于慈善捐赠的研究发现，在单独评估时，人们常出现数量忽略的问题，Desvousges 等(1992)很早就发现了人们对数量不敏感，一个经典例子就是，在单独评估时，人们给予两千、两万或者二十万只被石油污染困住的候鸟的捐赠金额没有什么差异。此外，受害者识别效应的研究也显示，人们会捐赠更多金钱给一位受害者，而不是人数更多的一群受害者(Kogut & Ritov, 2005)。其实，数量忽略问题和前文中提到的关于违背占优性原则的现象非常相似，正是因为人们对冰激凌的大小不够敏感、对棒球收藏卡片数量的不够敏感，所以才产生了不合理的决策。

根据广义可评估性理论，在这些情景中，人们知识较少或者缺乏内在参考系统(例如对于鸟类群体的数量)，又处于单独评估模式，总体处于可评估性较低的情况，因此对于数量和时间都比较不敏感。根据规范性解释，人们应该给数量较多的鸟群和人数较多的受害者捐款更多，但是由于人们对数量不敏感，做出的捐赠决策和人数上的实际需要不相符合。

如何增强我们对数量的可评估性，使得人们为更多人捐款时真的能捐助更多呢？Hsee, Zhang, Lu 和 Xu (2013)提出了一种新方式：单位请求(unit asking)。例如：某慈善网站在为 20 个低收入家庭的儿童募款，单位请求的方式要求捐赠者首先考虑，“如果你要帮助其中一个儿童，请问你愿意捐多少钱？”，然后再回答，“为 20 位儿童捐款，请问你愿意捐多少钱？”。

单位请求的方式先简单询问了一个“单位”的数值，创造了一种类似联合评估的比较，使得人们对较大的数量有了更准确的评估，从而提高了捐款的总量。因此，如果数量信息是影响人们明智决策的重要因素，但是普通人又不具备相应的知识时，采用单位请求这类助推方式，提升该因素的可评估性，形成像联合评价那样的效果，是一种简洁有效的做法。

3.4 消费选择

现实生活中，除了客观的决策结果(例如哪个理财方案最终收益最高)，人们也有许多决策的目标是让自己获得愉快的感受，也就是 Kahneman, Wakker 和 Sarin (1997)所提到的，让自身体验效用最大化。但是，在这个过程中，人们却不见得能准确选择。这是因为：人们预测或者购买时往往处于联合评估的模式，但实际生活或者使用时处于单独评估的模式，因此预测时人们容易高估体验时的价值敏感性。Hsee 和 Zhang (2004)将这种过分敏感的预测称为区别偏差 (distinction bias)。

首先，区别偏差会导致预测偏差。例如 Dunn, Wilson 和 Gilbert (2003)对大学生的调查发现：新生刚入学时往往认为宿舍的特征会影响自己的居住体验，因为刚申请宿舍时，大学生更接近联合评估模式；但是实际居住一段时间后调查发现：宿舍特征对于学生居住幸福感的影响很小，这是因为住在其中的学生更接近单独评估模式。

研究者发现，区别偏差导致的预测偏差有以下几种情形。第一，预测者高估了数量信息对体验者的影响。例如：Hsee 和 Zhang (2004)让大学生想象自己出版了一本诗集，并且校园书店正在销售此书。联合评估下，想象自己的诗集卖出了 0 本、80 本、160 本、240 本，对每种情况都评估自己的感受如何。单独评估组的参与者只看到了四种数量的其中一种，评估自己感受如何。在联合评估时，人们认为自己的感受呈现线性上升的趋势，四组之间差异显著。但是单独评估时，只要有诗集卖出，人们就比没有卖出诗集更加快乐，但是具体是卖出了多少本(80 本、160 本还是 240 本)对情绪就不再有影响。此外，Hsee 和 Zhang (2004)让参与者预测其他人阅读 10 个或者 25 个积极词汇(或者消极词汇)后情绪如何，预测者处于联合评估模式，认为词汇的效价(积极或消极)和词汇的数量(10 个还是 25 个词)都会影响阅读者的情绪。但是由于阅读者只阅读四种列表中的一种，处于单独评估模式，他们的情绪其实只受到词汇效价的影响，而不受到数量的影响。这与前文提到的评估模式和类别–数量信息的交互作用是相吻合的。

第二，预测者高估了比较信息对体验者的影响。例如：Hsee, Shen, Zhang, Chen 和 Zhang (2012)考察了天命注定和自由竞争两种分配方式下人们的感受如何。预测者认为自由竞争组的人们会感觉更公平；但是在实际体验者(只处于一种条件)中，天命注定组和自由竞争组的公平感并没有差异。体验者在单独评估的条件下，判断公平的参照信息较

少, 对公平的敏感程度并没有预测者想象的那么高。类似的, Cooney, Gilbert 和 Wilson (2016)也发现: 分配者认为分配过程的公平性很重要, 但是被分配者其实并没有那么重视过程。

由于人们在联合评估时会错误预测自身感受, 预测偏差也会导致实际的选择偏差, 使得人们去选择那些并不能使自己最愉快的选项。

第一类选择偏差表现在: 人们在选择时注重数量差异, 忽视类别差异, 但是在体验时更多受到类别差异的影响。例如: Hsee 和 Zhang (2004)让人们在两种任务间决策: 讲一个自己失败的故事并获得 15 克巧克力, 或者讲一个自己成功的故事并能获得 5 克巧克力。为了获得更大块的巧克力, 大部分人都愿意选择去讲失败的故事。但是单独体验这两种不同任务时, 讲成功故事获得少量巧克力的参与者更开心。因为巧克力的多少是数量信息, 体验者处于单独评估模式, 大块或者小块的巧克力影响不大; 但是讲述失败或成功故事带来的难过或愉快体验却有很大影响。

第二类选择偏差表现在: 在选择时, 个体更注重外在可比较的信息, 忽略了内在可评估的属性, 但在体验时, 个体却更多受到内在可评估属性的影响。例如: 前文提到的电子相框的选择(Yang et al., 2011), 联合评估时, 人们更多选择难看但分辨率高的相框, 但是, 让人们实际体验时, 反而是分辨率低但是好看的相框让人们体验更好, 因为单独摆放在家里时, 分辨率的差异是内在难以评估的, 外观美丑却是内在容易评估的。

因此, 可以看到, 联合评估固然能促进人们更为耐心、更合理地进行慈善捐赠, 但这不代表联合评估总是能帮助人们做出好的决策。要减弱区别偏差的影响, 使得消费者选择到使自己体验效用最大化的选项, 消费者需要在预测或选择阶段就采用单独评估模式, 最直接的方式就是, 让选项单独出现在消费者面前, 但是这种做法在现实中实行难度较高, 其他更可行的方式还有两种。

第一, “半单独评估”法, 也就是说, 消费者可以每次单独地观察一个选项, 间隔一段时间后再去评估其他选项。Yang 等(2011)对比了“半单独评估”和联合评估的决策差异。在联合评估下, 参与者多数选择外观丑陋但分辨率高的相框, 但是他们的体验却并不美好; 在半单独评估模式下, 参与者先观看一个相框, 间隔 10 分钟后, 再去观看另一个相框, 然后再作出选择, 这种情况下, 更多的参与者选择了分辨率低但是外观漂亮的相框, 也就是令人体验更愉悦的相框。因此, 这种方式可以降低区别偏差带来的选择偏差。

第二, 弱化技术参数。同样是上文中的选相框实验, 如果研究者不告知参与者关于分辨率的信息, 大部分参与者在仔细观察后会选择外观漂亮的相框(Yang et al., 2011), 即在体验阶段令消费者更愉快的那个选项。当然, 许多商家希望向消费者提供更多、更准确的信息, 但是, 如果一个消费决策的最终目的是拥有更愉快的感受, 那么弱化技术参数等数字的影响, 让消费者听从自己的体验和实际感受, 可能是一个更好的选择。

3.5　商业定价

区别偏差并不仅仅会导致个体做出不符合自身体验效用最大化的选择, 甚至有可能导致商业决策中的错误, 使得决策者无法实现经济利润的最大化。

Shen, Hsee, Wu 和 Tsai (2012)提出, 定价者一般处于联合评估模式, 会考虑到多种参考产品的价格来确定产品定价, 但是购买者往往只看到这个产品的价格, 处于单独评估模式。当购买者对其他参考产品的价格有所了解时, 一般不存在定价偏差的问题, 因为定价者和购买者都处于联合评估模式。但是, 当某种产品比较独特或者陌生时, 就有可能存在定价偏差——定价者处于联合评估模式, 购买者处于单独评估模式, 定价者很可能高估购买者的价格敏感性, 定价偏低, 未能实现利润最大化的目标。

为了验证定价者和购买者处于不同的评估模式, Shen 等(2012)让具有多年工作经验的 MBA 学生扮演定价者和购买者。定价者需要对一个可以识别情绪的新型相机定价, 目标是最大化营业额(即产品售价乘以售出数量)。在定价结束后, 报告自己的思考过程: 是仅仅思考了一个最有利润的价格(单独评估)还是思考了一系列价格后选择了其一(联合评估)。购买者被告知该相机以 3000 人民币的价格出售, 请决定是否愿意购买, 随后报告自己的思考过程, 是仅仅思考了在该价格下是否愿意购买(单独评估)还是考虑了其他可能的价格(联合评估)。结果发现: 定价者的确比购买者更多处于联合评估模式, 而且, 联合评估模式下的定价者制定出的价格显著低于单独评估模式下的定价者。

为了进一步说明定价者处于联合评估模式下时会产生定价误差, Shen 等(2012)招募了另一批

MBA 学生进行一个定价购买游戏。对定价者而言，他们要为一个业余摄影家的艺术照片定价，以获得最高利润。为简化任务，研究者免费给他们提供照片，照片成本为零；其次，他们有 50 份照片，将要以同样的价格出售给 50 名和他们类似的 MBA，50 位购买者每人单独决定自己是否购买 1 张；未售出的照片将作废，价值为零。定价者被随机分为控制组、单独评估组、联合评估组。结果发现：单独评估组的定价者制定的价格(中位数 50 元)要显著高于联合评估组和控制组(这两组中位数都是 20 元)。根据定价者的定价结果，研究者设置了高价组(每张照片 50 元)和低价组(每张照片 20 元)，购买者被随机分配到其中一种情况，决定自己是否购买一张艺术照片。结果发现：对于购买者而言，高价还是低价并没有影响购买者的比例。综合而言，单独评估的定价者制定的价格才是利润最大化的价格。

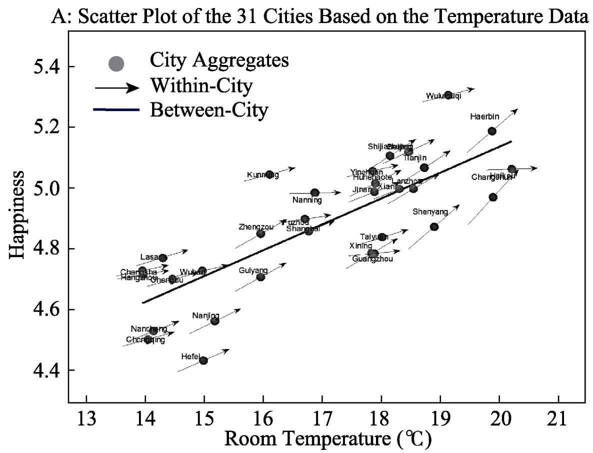
因此，当产品比较独特或者陌生时，定价者采用联合评估模式下制定的价格可能会有所偏差，定价者很可能高估处于单独评估模式下的购买者的价格敏感性，定价偏低，未能实现利润最大化的目标。在这时候，商业机构应该考虑重新审视自己的定价策略，让决策者们在单独评估的模式下思考，或者在对消费者进行市场调研时采取单独评估的调查策略，更有可能获取合理的参考定价。

3.6 幸福感

正如前文所述，很多时候，人类不能准确地判断什么才能给自己带来幸福快乐。我们以为重要的居住环境(Dunn et al., 2003)、分配方式(Hsee et al., 2012)、经济收入(Kahneman & Deaton, 2010)等因素可能并不像我们想象地那样带来愉快的体验和感受。人们的幸福到底是来源于哪些因素，是社会比较让我们更满足吗？通过一项跨城市的幸福感调查，Hsee, Yang, Li 和 Shen (2009)分析了这一问题。

研究者对中国 31 个城市的居民进行了一次电话调查，调查者询问了参与人的室内温度(调查在冬季进行)、拥有珠宝的价值以及对这两个方面的满意程度。结果发现：珠宝的价值对于幸福感的影响只存在相对效应(本城内珠宝越贵越幸福)，而不存在绝对效应(珠宝的绝对价值对幸福感没有影响)；但是室内温度对于幸福感的影响却存在绝对效应，冬季室内温度的绝对高低对于幸福感有显著影响，室温越高，幸福感越强，跨城市的效应仍然显著(如图 4 所示)。

可能的原因是：人们对珠宝没有内在的评估系统，进行判断时，城市内的社会比较信息可得性高，在城市内部接近联合评估模式，因此珠宝的相对价值对幸福感有影响；但是一旦扩大到城市外，社会比较信息可得性较低，在城市之间接近单独评估模式，珠宝价值就影响很小了。相反，人类对于舒适温度有着内在的参照系统，即使人们处于不同城市，没有其他地区的室温进行社会比较，接近单独评估模式，室内温度仍然是可评估性较高的属性，在跨城市的层面上，温度对幸福感仍有显著影响。



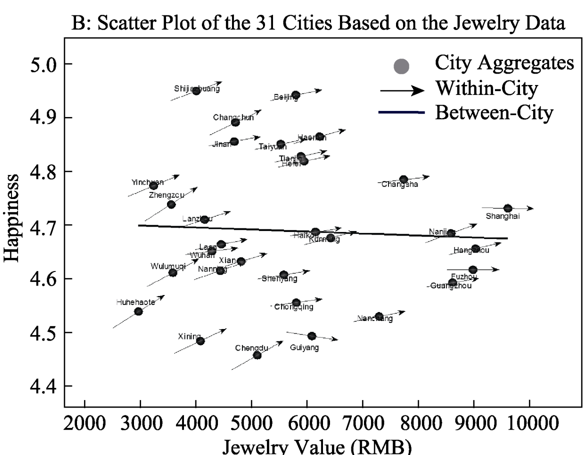


图 4 室内温度和珠宝对幸福感的影响
(资料来源: Hsee et al, 2009)

也就是说：内在可评估性高的因素对于幸福感的影响是绝对的，不依赖于社会比较信息或者其他外部参考信息。比如 关于食物品尝的研究发现：无论旁边放的是更美味的巧克力还是更难吃的沙丁鱼，人们在品尝薯条时的实际感受并没有什么差异，因为我们沉浸在薯条的味道中时，并不会过多思考其他食品(Gilbert & Wilson, 2009)。

Tu 和 Hsee (2016)还提出内在可评估性高的因素也更少受到享乐适应的影响。内在可评估性低的因素提升后，过一段时间，我们就产生了享乐适应，这种变化就不再影响我们的幸福感了。例如：从 1

克拉的钻石换成 2 克拉的，几个月后，当初的幸福感就消退了，因为变化后时间越久，个体的评估模式越接近单独评估，没有 1 克拉的鲜明比较，2 克拉钻石的可评估性降低了。但是，内在可评估性高的因素却不会如此，例如：室内温度、洗澡水温度等方面的变化，即使没有了过去的寒冷作为对照，我们也仍然对温度是否舒适非常敏感。因此，在提升幸福感的决策中，更为智慧的做法是去提升那些内在可感知的因素，例如添置暖风空调、品尝美味食物等比购买珠宝能带来更加长久的幸福感。

综上，单独评估和联合评估能深刻影响人们的风险偏好、跨期选择、消费体验等多方面判断和决策。政策制定者可以基于决策情景和决策目标的差异，采用不同的选择结构来助推人们做出适合的选择。当决策有清晰的外部目标，需要人们对风险更为中性、更有耐心可延迟满足、对数量信息的变化更为敏感时，宜采用联合评估。即使在决策情境难以实现联合评估时，可以设立参照点(例如单位请求法)，或者将数量信息转化为类别信息，来助推人们的决策。反之，当决策目标是提升个体的体验幸福时，宜采用单独评估，可以通过半单独评估、减少数量信息等方式，来避免人们进行直接的比较，获得更愉快的体验。

4　结论和展望

正如渴望成功的开发商不会选择众所周知、开发完善的地段，而会寻找被众人所忽视的潜力地段。如果一位学者希望做出富有影响力的研究，那么他/她也不应再进入那些研究已经非常充足的红海领域，而应去寻找那些富有潜力的新领域。我们认为，评估模式(单独评估与联合评估)的研究正是一支潜力股，存在多种的探索方向。

相比于其他的变量，单独和联合评估这一变量在现实中可以简洁有效地进行操纵，方便进行现场实验。Newman, Howlett 和 Burton (2016)关于标签的研究给了一个很好的演示：在单独评估的状态下，除了客观的线索，增加一些主观评估的标签能帮助消费者更流畅的决策，甚至做出更好的决策。例如在速食食品上，根据其营养成分表，研究者增加了一个主观标签“健康星级”。相比于没有健康星级标签的组，呈现了健康星级组的参与者选择健康食品的概率从 48%提升至了 74%。通过组织选项的呈现方式(单独或者联合呈现)，单独评估和联合评估非常容易进行操纵。如果研究者有志于让决策研究从实验室走到现实中去，考虑单独和联合评估这一领域可能是一个较为可行的选择。

在本文中，我们回顾了单独评估和联合评估带来的偏好反转现象，在许多重要的领域中，无论是人事招聘、消费选择、政策评估、道德判断、风险和跨期决策等，评估模式都影响了人们的判断和决策。广义可评估性理论对 JE/SE 偏好反转的成因进行了解释。评估模式也可以在风险和跨期决策、慈善捐赠、个体消费选择、商业定价等各个领域助推明智的决策。我们认为：两种评估模式是一种简洁并且有影响力的因素，而且这方面的研究还充满了新的机遇和空间，未来研究者会不断发现新的现象，探索更多的现实应用。

致谢：衷心感谢陈佳、李希琳、莫子川、戚知之、沈靓、王可、阮博文、张艽对本文的建议和修改。

参　考　文　献

Bazerman, M. H., Gino, F., Shu, L. L., & Tsay, C.-J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emotion Review, 3*(3), 290–292.

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly, 37*(2), 220–240.

Chatterjee, S., Heath, T. B., & Min, J. H. (2009). The susceptibility of mental accounting principles to evaluation mode effects. *Journal of Behavioral Decision Making, 22*(2), 120−137.

Cooney, G., Gilbert, D. T., & Wilson, T. D. (2016). When fairness matters less than we expect. *Proceedings of the National Academy of Sciences of the United States of America, 113*(40), 11168–11171.

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Boyle, K. J., Hudson, S. P., & Wilson, K. N. (1992). *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy*. Research Triangle Park, NC: RTI Press.

Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2003). Location, Location, Location: The misprediction of satisfaction in housing lotteries. *Personality and Social Psychology Bulletin, 29*(11), 1421–1432.

Evers, E. R. K., Inbar, Y., Blanken, I., & Oosterwijk, L. D. (2017). When do people prefer carrots to sticks? A robust “matching effect” in policy evaluation. *Management Science, 63*(12), 4261–4276.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics, 110*(3), 585–603.

Fredrick, S., Loewenstein, G., & O’Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*, 351–401.

Gilbert, D. T., & Wilson, T. D. (2009). Why the brain talks to itself: Sources of error in emotional prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1335–1341.

Gino, F., Moore, D. N., & Bazerman, M. H. (2008). No harm, no foul: The outcome bias in ethical judgments. Harvard Business School Working Paper, No. 08-080.
Gustafson, C. R., Lybbert, T. J., & Sumner, D. A. (2016). Consumer knowledge affects valuation of product attributes: Experimental results for wine. *Journal of Behavioral and Experimental Economics, 65*, 85–94.
Hsee, C. K. (1993). When trend of monetary outcomes matters: Separate versus joint evaluation and judgment of feelings versus choice. Working paper.
Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*(3), 247–257.
Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making, 11*(2), 107–121.
Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86*(5), 680–695.
Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*(4), 343–355.
Hsee, C. K., Shen, L. X., Zhang, S., Chen, J. Q., & Zhang, L. (2012). Fate or fight: Exploring the hedonic costs of competition. *Organizational Behavior and Human Decision Processes, 119*(2), 177–186.
Hsee, C. K., Yang, Y., Li, N. H., & Shen, L. X. (2009). Wealth, warmth, and well-being: Whether happiness is relative or absolute depends on whether it is about money, acquisition, or consumption. *Journal of Marketing Research, 46*(3), 396–409.
Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological Science, 24*(9), 1801–1808.
Hsee, C. K., Zhang, J., Wang, L. Y., & Zhang, S. (2013). Magnitude, time, and risk differ similarly between joint and single evaluations. *Journal of Consumer Research, 40*(1), 172–184.
Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America, 107*(38), 16489–16493.
Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty, 9*(1), 5–37.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.
Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics, 112*(2), 375–405.
Kogut, T., & Ritov, I. (2005). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes, 97*(2), 106–116.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443–478.
List, J. A. (2002). Preference reversals of a different kind: The “more is less” phenomenon. *American Economic Review, 92*(5), 1636–1643.
Newman, C. L., Howlett, E., & Burton, S. (2016). Effects of objective and evaluative front-of-package cues on food evaluation and choice: The moderating influence of comparative and noncomparative processing contexts. *Journal of Consumer Research, 42*(5), 749–766.
Okada, E. M. (2005). Justification effects on consumer choice of hedonic and utilitarian goods. *Journal of Marketing Research, 42*(1), 43–53.
Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes, 109*(2), 134–141.
Powdthavee, N., Riyanto Y. E., & Knetsc, J. K. (2017). Economists’ judgments of publication lists with lower ranked journals included: Evidence from a survey experiment. Working paper.
Shaffer, V. A., & Arkes, H. R. (2009). Preference reversals in evaluations of cash versus non-cash incentives. *Journal of Economic Psychology, 30*(6), 859–872.
Shen, L. X., Hsee, C. K., Wu, Q. S., & Tsai, C. I. (2012). Overpredicting and underprofiting in pricing decisions: Evaluability in pricing decisions. *Journal of Behavioral Decision Making, 25*(5), 512–521.
Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General, 143*(3), 1127–1143.
Tu, Y. P., & Hsee, C. K. (2016). Consumer happiness derived from inherent preferences versus learned preferences. *Current Opinion in Psychology, 10*, 83–88.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior (60th Anniversary Commemorative Edition)*. Princeton: Princeton University Press.
Yang, A. X., Hsee, C. K., Liu, Y., & Zhang, L. (2011). The supremacy of singular subjectivity: Improving decision quality by removing objective specifications and direct comparisons. *Journal of Consumer Psychology, 21*(4), 393–404.
Zhang, J. (2015). Joint versus separate modes of evaluation. In G. Keren & G. Wu (Eds.), *The wiley blackwell handbook of judgment and decision making* (pp. 211–238). Portsmouth: John Wiley & Sons, Ltd.
Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2004). “Is 28% good or bad?” Evaluability and preference reversals in health care decisions. *Medical Decision Making, 24*(2), 142–148.

Joint evaluation versus single evaluation: A field full of potentials

LU Xi[1,2]; HSEE, Christopher K.[3]

([1] Guanghua School of Management, Peking University, Beijing 100871, China)

([2] College of Economics and Management, China Agricultural University, Beijing 100083, China)

([3] Booth School of Business, The University of Chicago, Chicago IL 60637, United States)

Abstract

Evaluation mode is an important yet under-studied aspect of decision-making. All decisions are made in one of two evaluation modes or some mix of the two. One is joint evaluation (JE), in which different options are presented together and evaluated simultaneously. The other is single evaluation (SE), in which different options are presented separately, and each option is evaluated in isolation.

This article reviews existing literature on how JE and SE can lead to reversed preferences, with one option favored in JE and another option favored in SE, in domains including hiring, consumer choice, moral judgments and healthcare decisions. For example, in JE, a plain-looking but experienced job candidate would be favored over a good-looking but less-experienced job candidate, but in SE, the good-looking but less experienced candidate would be favored.

We also review existing research on the “less is better” phenomenon–that in SE (but not in JE), a normatively less valuable option is judged more favorably than its more valuable alternative. For example, in SE (but not in JE), 7 oz. of ice cream served in a 5-oz. cup is valued more favorably than 8 oz. of ice cream served in a 10-oz. cup.

To interpret JE/SE preference reversals and the less-is-better effect, several explanations have been proposed, such as a want/should conflict, a within-category versus between-category comparison, the option as information model, and the dual-process model. The present review focuses on the general evaluability theory (GET), which ascribes JE/SE preference reversals and the less-is-better effect to attribute evaluability. According to GET, the evaluability of an attribute depends conjunctively on three factors: evaluation mode (JE versus SE), knowledge of the decision-maker about the attribute, and the inherent/learned nature of the attribute. Not only can GET explain JE/SE preference reversals and the less-is-better effect, it also explains many other effects, such as scope neglect, differences in risk preference between JE and SE, differences in time preference between JE and SE, misprediction of future hedonic experiences, and so on.

Understanding JE-SE differences and attribute evaluability also offers insight into how to improve decisions in contexts such as fundraising, pricing strategy, public service, and subjective well-being.

The review suggests that evaluation mode (JE versus SE) is an important independent variable that influences a wide range of decisions in counterintuitive ways and thereby provides opportunities to design choice-architecture-based nudges to improve decisions. Yet despite its importance, evaluation mode has not been well-studied and therefore is a fertile ground for researchers to explore and develop.

Key words　evaluation mode; preference reversal; general evaluability theory; dual-process model; nudge